HW 5

Zoe Werner

11/11/2023

The COMPAS algorithm provides a statistical evaluation of the likelihood of an incarcerated person recommitting crimes. It was created as an asset for judges, providing them with a statistical opinion on each probation case, but creator (Northpointe) was sued by several incarcerated people that claimed the algorithm was unfairly bias towards them due to racial prejudice. While the COMPAS algorithm was created with good intentions to act as a statistical reference tool for aiding judges in probation decisions, its predictions included bias that resulted in unfair prejudice and should not be used in court proceedings.

Beginning with the statistical evidence of the COMPAS algorithm, the predictions violated the criteria for evaluating statistical methods of fairness. It did not pass the independence criteria requiring statistical parity and a lack of disparate impact. because it violated statistical parity and had disparate impact. The statistical parity determines whether each group (in this case, racial groups) has similar probability of being included in each classification prediction group. The disparate impact determines the probability of each group receiving a positive classification and evaluates the difference between then. The rate is compared to 1 - epsilon, which in our case is 0.2, the legal precedent. Since the COMPAS disparate impact rate is less then 0.8, the algorithm has disparate impact. Both the presence of disparate impact and the violation of statistical parity results in a failure to meet the independence criteria of fairness bias, proving that black convicts had a larger chance of being classified as re-offenders than white convicts. In addition, the separation criteria based on the equalized odds test was violated. The equalized odds test ensures that accuracy rates are equal across difference groups and in this case, equal among white and black convicts. In order to pass the equalized odds test, both the true positive test and the false negative test must not be violated. The algorithm did not violated the true positive rate, which resulted in 0.1974, which is lower then 0.2. However, the difference ratio of true positive rates for white and black convicts was above 0.2, resulting in a violation. This proved that white convicts were more likely to receive a true positive prediction than black convicts.

In addition to the statistical evidence of bias, the COMPAS algorithm also violated multiple philosophical frameworks. COMPAS fails to take into account any changes that a convict may have made during their incarceration. Since the algorithm is based solely on data and probability, it fails to consider rehabilitation efforts or any positive actions/behavior since being incarcerated. It doesn't have access to information that we do as humans, such as human compassion or the ability to observe changes in a convicts demeanor or behavior. In turn, using the COMPAS algorithm violates the philosophical framework of virtue ethics due to its lack of compassion, fairness, and justice. While one may argue that the algorithm helps support justice, the unfair denial of probation after a convicts time has been served is not justice. Justice is impartial, objective, rational-decision making and is characterized by fairness and equality. A cardinal value of the United States Justice System is the consideration that a person is innocent until proven guilty. While convicts seeking probation aren't innocent, they should not be labeled as a re-offender when they haven't proven themselves to be. There are exceptions, such as murder and other violent crimes, but this should be up to a human judge, not an algorithm that does not have the capacity to act justly. An algorithm cannot feel compassion since it is not human and when examining a convicts record and behavior while incarcerated, it is sometimes necessary to have the capacity to feel compassion. For example, if someone was involved in accidental manslaughter and had a shoplifting record from their youth, should they be denied bail simply because it was a second offense? This type of decision should be made by a human judge who can attempt to assess the convicts intentions and actions with the possibility of being compassionate towards their situation. While a decision should never be made strictly based on compassion, it is an important consideration that

should be included in the judicial process when necessary. Finally, as we evaluated in the statistical section, the algorithm does not achieve the virtue of fairness. It has statistical evidence of unfair bias based on race, which is certainly not achieving fairness.

Finally, the COMPAS algorithm violates the philosophical framework of Consequentialism, which states that an actions justifiability is completely based on its outcome, and specifically violates the subset Utilitarianism, which states that a rule permitting action should result in net pleasure, so if the costs outweigh the benefits, it is not morally permissible. Since the algorithm is only meant to supplement a judge's decision, it doesn't provide an overwhelming pleasure to the judge. They are still expected to come to a conclusion on their own and will just be using the algorithm results as an additional piece of evidence. On the contrary, the costs of bias resulting in unjust probation denials is quite abundant. The convict is forced to remain in jail and lose more time they could be spending with their family, friends, working at a job, or pursuing education. Therefore, the pain caused by unfair, incorrect re-offender classifications based on the COMPAS algorithm outweighs the minimal benefit of supplementary evidence for the judge.

In conclusion, the COMPAS algorithm violates the criteria of statistical methods of fairness, the philosophical framework of Utilitarianism, and lacks the capacity to include virtue ethics in its decision-making. Therefore, judges should not be permitted to utilize COMPAS in their probation hearings, instead relying on their personal ruling of each case-by-case situation.